Frontier of Artificial Network

Department of Data Science
香港城市大學
City University of Hong Kong

Bridging Scene Understanding and Motion Generation in Robot Manipulation

Invited Speaker
Zeyu Zhang
Beijing Institute for General Artificial Intelligence

Date: Feb 5, 2025 (Thu)
Time: 14:00-15:00 (HKT)
Zoom Meeting: 801 137 0362

Biography

Zeyu Zhang is a Senior Researcher at the Beijing Institute for General Artificial Intelligence (BIGAI) and serves as a doctoral co-advisor in the Tong Program in collaboration with Peking University and Huazhong University of Science and Technology. He received his Ph.D. in Computer Science from the University of California, Los Angeles (UCLA), under the supervision of Professor Song-Chun Zhu. He is a recipient of the Beijing Overseas Young Talent Award. He posits that human intelligence is grounded in fundamental representations and a unified cognitive architecture. His research seeks to uncover and computationally realize this architecture to enable higher levels of autonomy and intelligence in robotics by bridging how robots perceive the world, reason over abstract structures, and act in complex environments. His work has been published in leading journals and conferences in the field, including IJCV, TPAMI, Engineering, T-RO, R-AL, ICRA, IROS, and ICCV, with over 20 publications to date. He has long served as a reviewer for top-tier international conferences and journals.

Abstract

In this talk, I will present a series of our works on robot manipulation in 3D scenes that aim to discover a unified representation bridging perception and action. I will begin with how we model the embodiment of the robot and the manipulated object through a unified kinematic representation using a Virtual Kinematic Chain. I will then introduce our function-oriented scene reconstruction that enables robots to perceive environments as actionable structures for manipulation. Next, I will show how we jointly infer perception, task plans, and motion generation in an object cutting example by combining neural signals with symbolic recursive reasoning. Finally, I will present our recent work on mobile manipulation that generalizes these ideas to coordinated whole-body behaviors across diverse 3D scenes.